

## Durham Research Online

---

### Deposited in DRO:

16 March 2018

### Version of attached file:

Accepted Version

### Peer-review status of attached file:

Peer-reviewed

### Citation for published item:

Kuhn, Patrick M. and Vivyan, Nick (2018) 'Reducing turnout misreporting in online surveys.', *Public opinion quarterly*, 82 (2). pp. 300-321.

### Further information on publisher's website:

<https://doi.org/10.1093/poq/nfy017>

### Publisher's copyright statement:

This is a pre-copyedited, author-produced version of an article accepted for publication in *Public Opinion Quarterly* following peer review. The version of record Kuhn, Patrick M. Vivyan, Nick (2018). *Reducing Turnout Misreporting in Online Surveys*. *Public Opinion Quarterly*, 82(2): 300-321 is available online at: <https://doi.org/10.1093/poq/nfy017>.

### Additional information:

---

### Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

## **Reducing Turnout Misreporting in Online Surveys**

Forthcoming in *Public Opinion Quarterly*

### **Author list**

PATRICK M. KUHN: Durham University, School of Government and International Affairs,  
Al Qasimi Building, Elvet Hill Road, Durham DH1 3TU, UK; Phone: +44 191 334 7204;  
Email: p.m.kuhn@durham.ac.uk.

NICK VIVYAN: Durham University, School of Government and International Affairs, Al  
Qasimi Building, Elvet Hill Road, Durham DH1 3TU, UK; Phone: +44 191 334 5881; Email:  
nick.vivyan@durham.ac.uk. (Corresponding Author)

## **Author note**

PATRICK M. KUHN is an assistant professor in the School of Government and International Affairs, Durham University, Durham, UK. NICK VIVYAN is an associate professor in the School of Government and International Affairs, Durham University, Durham, UK. The authors thank participants at the North East Research Development (NERD) Workshop and the 2016 annual meeting of the European Political Science Association in Brussels, as well as the editors and three anonymous reviewers for helpful comments. This work was supported solely by internal funding to Patrick Kuhn from the School of Government and International Affairs, Durham University. \*Address correspondence to Nick Vivyan, Durham University, School of Government and International Affairs, Al Qasimi Building, Elvet Hill Road, Durham DH1 3TU, UK; e-mail: [nick.vivyan@durham.ac.uk](mailto:nick.vivyan@durham.ac.uk).

## **Abstract**

Assessing individual-level theories of electoral participation requires survey-based measures of turnout. Yet due to a combination of sampling problems and respondent misreporting, postelection surveys routinely overestimate turnout, often by large margins. Using an online survey experiment fielded after the 2015 British General Election, we implement three alternative survey questions aimed at correcting for turnout misreporting and test them against a standard direct turnout question used in postelection studies. Comparing estimated to actual turnout rates, we find that while all question designs overestimate aggregate turnout, the item-count technique alleviates the misreporting problem substantially, whereas a direct turnout question with additional face-saving options and a crosswise model design help little or not at all. Also, regression models of turnout estimated using the item-count measure yield substantively similar inferences regarding the correlates of electoral participation to models estimated using 'gold-standard' validated vote measures. These findings stand in contrast to those suggesting item-count techniques do not help with misreporting in an online setting and are particularly relevant given the increasing use of online surveys in election studies.

## Introduction

Self-reported turnout rates in postelection surveys often considerably exceed official rates.<sup>1</sup> This phenomenon of 'vote overreporting' (e.g., Bernstein, Chadha, and Montjoy 2001; McDonald 2003) represents a major challenge for election research, raising questions about the validity of turnout models estimated using survey data (e.g., Brehm 1993; Bernstein, Chadha, and Montjoy 2001; Cassel 2003; Karp and Brockington 2005; Jones 2008).

While vote overreporting is attributable in part to sampling and survey nonresponse biases (e.g., Brehm 1993; Jackman 1999; Voogt and Saris 2003), much previous research focuses on the tendency of survey respondents - particularly those who did not vote - to *misreport* their turnout (Presser 1990; Abelson, Loftus, and Greenwald 1992; Holtgraves, Eck, and Lasky 1997; Belli, Traugott, Young and McGonagle 1999; Belli, Moore, and VanHoewyk 2006; Holbrook and Krosnick 2010b; Hanmer, Banks, and White 2014; Persson and Solevid 2014; Zeglovits and Kritzing 2014; Thomas, Johan, Kritzing, Plescia and Zeglovitz 2017).

This paper investigates whether misreporting can be alleviated by different sensitive survey techniques designed to reduce social desirability pressures arising from

---

<sup>1</sup> The average difference between survey and official turnout rate across 150 Comparative Study of Electoral Systems (CSES) postelection surveys is around 12 percentage points (The Comparative Study of Electoral Systems 2017).

turnout-related questions. In particular, we examine the crosswise model (CM) and the item-count technique (ICT). Whereas these approaches had been of limited use to scholars estimating multivariate models of turnout, recent methodological advances (Blair and Imai 2010; Imai 2011; Blair and Imai 2012; Jann, Jerke, and Krumpal 2012; Blair, Imai, and Zhou 2015) have made estimating such models relatively straightforward. In an online survey experiment fielded shortly after the 2015 UK general election, we design new CM and ICT turnout questions and test them against a standard direct turnout question and a direct question with face-saving response options.

Our findings show that while all question designs overestimate aggregate national turnout, ICT yields more accurate estimates compared to the standard direct question, whereas the face-saving design and CM improve accuracy little or not at all. Also, regression models of turnout estimated using ICT measures yield inferences regarding the correlates of electoral participation that are more consistent with those from models estimated using 'gold standard' validated vote measures.

In contrast to recent studies that cast doubt on the suitability of ICT questions for reducing turnout misreporting in online surveys (Holbrook and Krosnick 2010b; Thomas et al. 2017), we show that ICT questions designed following current best practice appear to substantially reduce turnout misreporting in an online survey. Our results suggest that earlier mixed findings regarding ICT's effectiveness could be due to the particular ICT designs used in those studies.

## **Turnout as a sensitive topic**

Existing research has sought to alleviate turnout misreporting in a number of ways. One approach is to disregard self-reports and instead measure respondent turnout using official records. Such 'vote validation' exercises have been undertaken in several national election studies (e.g., in Sweden, New Zealand, Norway, the U.K. and - until 1990 - the US).

Although often considered the gold standard in dealing with misreporting, the vote-validation approach in the U.S. context has raised doubts, with Berent, Krosnick, and Lupia (2016) showing that matching errors artificially drive down "validated" turnout rates. While it is an open question to what extent matching errors are an issue outside the U.S. context, vote validation has two additional downsides that limit its utility as a general solution for turnout misreporting. First, in many countries official records of who has voted in an election are not available. Second, these records, when available, are often decentralized, making validation a time-consuming and expensive undertaking.

Another set of approaches for dealing with turnout misreporting focus on alleviating social desirability bias (for overviews see Tourangeau and Yan 2007; Holbrook and Krosnick 2010b). Voting is an admired and highly valued civic behavior (Holbrook, Green, and Krosnick 2003; Karp and Brockington 2005; Bryan, Walton, Rogers and

Dweck 2011), creating incentives for nonvoters to deliberately or unconsciously misreport when asked about their electoral participation.

Starting from this premise, some suggest that misreporting can be alleviated via appropriate choice of survey *mode*, with respondents more willing to report sensitive information in self- rather than interviewer-administered surveys (Hochstim 1967). Although Holbrook and Krosnick (2010b) find that turnout misreporting is reduced in self-administered online surveys compared to interviewer-administered telephone surveys, a systematic review of over 100 postelection surveys found no significant difference in turnout misreporting across survey modes (Selb and Munzert 2013). Reviewing studies on a variety of sensitive topics, Tourangeau and Yan (2007, 878) conclude that "even when the questions are self-administered,... many respondents still misreport."

If choice of survey mode alone cannot resolve the misreporting problem, can we *design* turnout questions that do? One design-based approach for reducing misreporting is the "bogus pipeline" (Jones and Sigall 1971; Roese and Jamieson 1993), where the interviewer informs the respondent that their answer to the sensitive question will be verified against official records, thus increasing the respondent's motivation to tell the truth (assuming being caught lying is more embarrassing than admitting to the sensitive behavior). Hanmer, Banks, and White (2014) find that this approach significantly reduces turnout misreporting. However, provided researchers do not want to mislead survey respondents, the applicability of the bogus pipeline is limited, since it



necessitates vote validation for at least some respondents, which is costly and sometimes impossible.

A simple alternative design-based approach is to combine 'forgiving' question wording (Fowler 1995), which attempts to normalize nonvoting in the question preamble, with the provision of answer options that permit the respondent to admit nonvoting in a 'face-saving' manner. Although turnout misreporting is unaffected by 'forgiving' wording<sup>2</sup> (Abelson, Loftus, and Greenwald 1992; Holtgraves, Eck, and Lasky 1997; Persson and Solevid 2014) and only moderately reduced by 'face-saving' answer options (Belli et al. 1999; Belli et al. 2006; Persson and Solevid 2014; Zeglovits and Kritzinger 2014), many election studies incorporate one or both of these features in their turnout questions. We therefore include such turnout question designs as comparators in our experiments.

Other design-based approaches to the misreporting problem involve *indirect* questions, which aim to reduce social desirability pressures by protecting privacy such that survey researchers are unable to infer individual respondents' answers to the sensitive item. The well-known randomized response technique ensures this using a randomization device: Warner (1965), for example, asks respondents to truthfully state either whether they *do* bear the sensitive trait of interest, or whether they *do not* bear

---

<sup>2</sup> Other changes to the preamble of a turnout question aimed at increasing truthful reporting, such as asking for polling station location, were equally unsuccessful (Presser 1990).

the sensitive trait of interest, based on the outcome of a whirl of a spinner unobserved by the interviewer. The researcher is thus unaware of which question an individual respondent is answering, but can estimate the rate of the sensitive behavior in the sample because she knows the probability with which respondents answer each item. Research suggests that this design fails to reduce turnout misreporting (Locander, Sudman, and Bradburn 1976; Holbrook and Krosnick 2010a) and raises concerns about its practicality: In telephone and self-administered surveys, it is difficult to ensure that respondents have a randomization device to hand and that they appropriately employ it (Holbrook and Krosnick 2010a).<sup>3</sup>

Recognizing these practical limitations, researchers have developed variants of the randomized response technique that do not require randomization devices. One recent example is the crosswise model (CM) (Yu, Tian, and Tang 2008; Tan, Tian, and Tang 2009) where respondents are asked two yes/no questions - a nonsensitive question where the population distribution of true responses is known, and the sensitive question of substantive interest - and indicate only whether or not their answers to the questions are identical. Based on respondents' answers and the known distribution of answers to the nonsensitive item, researchers can again estimate the

---

<sup>3</sup> Rosenfeld, Imai, and Shapiro (2016) do find that a randomized response design appears to reduce misreporting of sensitive vote choices, but also find evidence of potential noncompliance in respondent implementation of the randomization device (796).

rate of the sensitive trait. CM has been shown to reduce misreporting on some sensitive topics (e.g., Coutts and Jann 2011; Jann, Jerke, and Krumpal 2012), but is as yet untested with regard to turnout.

A final example of indirect questioning is the item-count technique (ICT), or 'list experiment.' In this design, respondents are randomized into a control and treatment group. The control group receives a list of nonsensitive items, while the treatment group receives the same list plus the sensitive item. Respondents are asked to count the total number of listed items that satisfy a certain criteria rather than answering with regard to each individual listed item. The prevalence of the sensitive trait is estimated based on the difference in mean item counts across the two groups (Miller 1984; Droitcour, Caspar, Hubbard, Parsley, Visscher and Ezzati 1991). The ICT performance record is mixed, both with regard to turnout (e.g., Holbrook and Krosnick 2010b; Comşa and Postelnicu 2013; Thomas et al. 2017) and other sensitive survey items (e.g., Tourangeau and Yan 2007; Wolter and Laier 2014). This mixed success may reflect the challenges researchers face in creating valid lists of control items --challenges that have been addressed in a recent series of articles (Blair and Imai 2012; Glynn 2013; Aronow, Crawford and Green 2015). Below, we investigate whether an ICT question designed according to current best practice can reduce nonvoter misreporting in an online survey.

## **METHODS**

### **Experimental Design**

Our survey experiment was designed to test whether new ICT and CM turnout question designs are effective at reducing misreporting, relative to more standard direct turnout

questions with forgiving wording and face-saving response options. Our experiment was run online through YouGov across four survey waves in the aftermath of the UK general election on May 7th 2015 (see Appendix for further sampling details). To limit memory error concerns, fieldwork was conducted soon after the election, 8-15 June 2015, with a sample of 6,228 respondents from the British population. Appendix Table A.2 reports sample descriptives, showing that these are broadly in line with those from the British Election Study (BES) face-to-face postelection survey, a high-quality probability sample, and with census data.

### **Survey instruments**

Respondents were randomly assigned to one of four turnout questions.

*Direct question.* Our baseline turnout question is the direct question used by the BES, which already incorporates a 'forgiving' introduction. Respondents were asked: "Talking with people about the recent general election on May 7th, we have found that a lot of people didn't manage to vote. How about you, did you manage to vote in the general election?" Respondents could answer yes or no, or offer "don't know." The estimated aggregate turnout from this question is the (weighted or unweighted) proportion of respondents answering 'Yes'.

*Direct face-saving question.* This variant incorporates the preamble and question wording of the direct question, but response options are now those that Belli et al.(2006) propose for when data are collected within a few weeks of Election Day: "I did not vote in the general election"; "I thought about voting this time but didn't"; "I usually vote but didn't this time"; "I am sure I voted in the general election"; Don't know.

The second and third answer options allow respondents to report nonvoting in the election while also indicating having had some intent to vote or having voted on other occasions, and may therefore make it easier for nonvoters to admit not having voted. Aggregate turnout is estimated as the (weighted or unweighted) proportion of respondents giving the penultimate response.

*Crosswise model (CM).* Our CM question involves giving respondents the following question: “Talking with people about the recent general election on May 7th, we have found that a lot of people didn't manage to vote or were reluctant to say whether or not they had voted. In order to provide additional protection of your privacy, this question uses a method to keep your answer totally confidential, so that nobody can tell for sure whether you voted or not. Please read the instructions carefully before answering the question.

Two questions are asked below. Please think about how you would answer each question separately (either with Yes or No). After that *please indicate whether your answers to both questions are the same* (No to both questions or Yes to both questions) *or different* (Yes to one question and No to the other).” The two questions were “Is your mother's birthday in January, February or March?” and “Did you manage to vote in the general election?”

This follows Jann, Jerke, and Krumpal (2012) in asking about parental birthdays as the nonsensitive question, as this satisfies key criteria for CM effectiveness (Yu, Tian, and Tang 2008): The probability of an affirmative response is known, unequal to 0.5 and uncorrelated with true turnout. We calculate the probability that a respondent's mother

was born in January, February, or March based on Office of National Statistics data on the birth dates of British women, 1938-1983. The calculated probability is 25.2%.

So that respondents understand why they are being asked such a complex question, and consistent with Jann, Jerke, and Krumpal (2012), the preamble explicitly states that the question is designed to protect privacy.

Following Yu, Tian, and Tang (2008), the CM estimate of aggregate turnout is  $\hat{\pi}_{CM} = (r/n + p - 1)/(2p - 1)$ , where  $n$  is the total number of respondents,  $r$  is the number who report matching answers, and  $p$  is the known probability of an affirmative answer to the nonsensitive question. The standard error is

$$\hat{se}(\hat{\pi}_{CM}) = \sqrt{((r/n)(1 - r/n))/((n - 1)(2p - 1)^2)}.^4$$

*Item-count technique (ICT).* In the ICT design, respondents were asked: “The next question deals with the recent general election on May 7th. Here is a list of four (five) things that some people did and some people did not do during the election campaign or on Election Day. Please say how many of these things you did.” The list asked respondents whether they had: discussed the election with family and friends; voted in the election (sensitive item); criticised a politician on social media; avoided watching the

---

<sup>4</sup> For weighted CM estimates, we replace the term  $r/n$  with  $\sum y_i w_i$ , where  $y_i$  is a binary indicator of whether respondent  $i$  reports matching answers, and  $w_i$  denotes the survey weight for observation  $i$ . Weights are standardized so that  $\sum w_i = 1$ . We also replace  $n$  in the denominator of the standard error equation with effective sample size based on Kish's approximate formula (Kish 1965).

leaders debates; and put up a poster for a political party in my window or garden.

Respondents could provide an answer between 0 and 4, or say they did not know.

This design incorporates a number of recommendations from recent studies of ICT effectiveness. First, to avoid drawing undue attention to our sensitive item, each nonsensitive item relates to activities that respondents might engage in during election periods (Kuklinski, Cobb, and Gilens 1997; Aronow et al. 2015; Lax, Phillips, and Stollwerk 2016). This contrasts with existing ICT-based turnout questions, which include non-political behaviors in the control list (Holbrook and Krosnick 2010b; Comşa and Postelnicu 2013; Thomas et al. 2017) and which have had mixed success in reducing misreporting.

Second, we are careful to avoid ceiling and floor effects, which occur when a respondent in the treatment group engages in either all or none of the nonsensitive behaviors and therefore perceives that their answer to the sensitive item is no longer concealed from the researcher (Blair and Imai 2012; Glynn 2013). To minimize such effects, we include a 'low-cost' control activity that most respondents should have undertaken ("discussed the election with family and friends") and a 'high-cost' activity that few respondents should have undertaken ("put up a poster for a political party").

In addition to implementing these recommendations, the control list includes some 'norm-defiant' behaviors, such as "avoided watching the leaders debate" and "criticised a politician on social media." Our intent here is to reduce embarrassment at admitting nonvoting by signalling to respondents that it is recognized that some people do not like and/or do not engage with politics.

Unlike the CM design, and consistent with standard ICT designs for online surveys (e.g., Aronow et al. 2015; Lax, Phillips, and Stollwerk 2016), the preamble does not explicitly state that the question is designed to protect privacy.

Our ICT-based estimate of aggregate turnout is the difference in (weighted or unweighted) mean item counts comparing the control and treatment groups (Blair and Imai 2012). For the weighted estimate, standard errors were calculated using Taylor linearization in the 'survey' package (Lumley 2004) in R.

Tests reported in Supplementary Materials Section B reports diagnostics suggesting that this ICT design successfully minimizes ceiling and floor effects and satisfies other key identifying assumptions laid out in Blair and Imai (2012).

### **Randomization**

Respondents were randomly assigned to one of the four turnout questions described above. Due to its lower statistical efficiency, ICT received double weight in the randomization. Of the 6,228 respondents, 1,260 received the direct question, 1,153 the direct face-saving question, 2,581 the ICT question, and 1,234 the CM question.

Supplementary Materials Section A suggests randomization was successful.



## RESULTS

### Comparing turnout estimates

We begin our analysis by comparing headline turnout estimates. Figure 1 displays, for each survey technique, weighted and unweighted Britain-wide turnout estimates. Given the similarity between weighted and unweighted estimates, we focus on the former.<sup>5</sup>

[INSERT FIGURE 1 ABOUT HERE]

The standard direct technique performs poorly, yielding a turnout estimate of 91.2% [89.3%, 93%], 24.7 points higher than actual turnout.

In line with previous U.S. (Belli et al. 2006) and Austrian (Zeglovits and Kritzing 2014) studies, the face-saving question yields a modest improvement. It significantly reduces estimated turnout compared to the direct technique, but still performs poorly in absolute terms, estimating turnout at 86.6% [84.1%, 89%], 20.1 points higher than actual turnout.

CM performs worst of all the techniques we test, estimating turnout at 94.3% [88.4%, 100%], 27.9 points higher than actual turnout.

In contrast, while ICT is clearly less efficient (with a relatively wide confidence interval), it nevertheless yields a substantively and statistically significant improvement

---

<sup>5</sup> We use standard YouGov weights, generated by raking the sample to the population marginal distributions of age-group  $\times$  gender, social grade, newspaper readership, region and party identification.

in turnout estimate accuracy compared to all other techniques.<sup>6</sup> Though still 9.2 points higher than actual GB turnout, the ICT estimate of 75.7% [66.9%, 84.4%] represents a two-third reduction in error compared to the direct question estimate. Taking the difference between the ICT and direct turnout estimates in our data, one gets an implied misreporting rate of 15.5% [6.5%, 24.4%]. The confidence interval contains -- and is therefore consistent with -- the 10% rate of misreporting found by Rivers and Wells (2015), who validate the votes of a subset of YouGov respondents after the 2015 general election.

In sum, the face-saving and ICT questions yield aggregate turnout estimates that are, respectively, moderately and substantially more accurate than those from the direct question, while CM yields no improvement.<sup>7</sup> ICT, however, still overestimates

---

<sup>6</sup> Despite the slight overlap in confidence intervals for weighted estimates, the differences between the weighted ICT and face-saving estimates are statistically significant (weighted,  $z = 3.39$ ,  $p$ -value  $< 0.01$ ; un-weighted,  $z = 4.33$ ,  $p$ -value  $< 0.01$ ). Schenker and Gentleman (2001) show that overlapping confidence intervals do not necessarily imply non-significant differences. The differences between ICT and direct estimates are also significant (weighted,  $z = 2.34$ ,  $p$ -value = 0.019; un-weighted,  $z = 3.43$ ,  $p$ -value  $< 0.01$ ).

<sup>7</sup> Supplementary Materials Section C shows that question effects are consistent when each of the four survey waves is treated as a distinct replication of our experiment.

actual 2015 turnout, which may partly be because ICT does not correct all misreporting. It may also be partly explained by the fact that while YouGov samples from this period have been found to overestimate aggregate turnout due to *both* misreporting *and* oversampling of politically interested individuals who are more likely to vote (Rivers and Wells 2015), ICT tackles only misreporting.

Before probing the face-saving and ICT results using multivariate analysis, we pause to consider why the CM design failed. One possibility is that, faced with a somewhat unusual question and in the absence of a practice run, some respondents found the CM question unduly taxing and simply answered 'don't know'. If the propensity to do so is negatively correlated with turnout, this could explain why CM overestimates turnout. However, Table 1 casts doubt on this explanation, showing that the proportion of 'don't know' responses are not substantially higher for CM compared to other treatments.<sup>8</sup>

[INSERT TABLE 1 ABOUT HERE]

---

<sup>8</sup> The difference in the rate of 'don't know' responses between CM and other treatments is often statistically significant ( $z = -2.51$ ,  $p$ -value = 0.012 for CM vs. direct question;  $z = -3.06$ ,  $p$ -value < 0.01 for CM vs. face-saving question;  $z = -0.13$ ,  $p$ -value = 0.9 for CM vs. ICT control;  $z = -2.32$ ,  $p$ -value = 0.02 for CM vs. ICT sensitive). However, the maximum magnitude of any difference in 'don't know' rates is 2 percentage points.

A more plausible explanation for the disappointing performance of our CM lies in a combination of two features of this design. First, while such an unusual design necessitates an explanatory preamble, stating that it represents 'an additional protection of your privacy' may heighten the perceived sensitivity of the turnout question for respondents (Clifford and Jerit 2015). Second, in the absence of a run-through illustrating how the design preserves anonymity, respondents whose sensitivity was heightened by the preamble may distrust the design and become particularly susceptible to social desirability bias. This is consistent with Coutts and Jann (2011), who find that in an online setting, randomized response designs -- which share many characteristics with CM -- elicit relatively low levels of respondent trust. Solving this problem is not easy: Doing a CM run-through in online surveys is time-consuming and may frustrate respondents.<sup>9</sup>

### **Comparing turnout models**

The improvement in aggregate turnout estimates yielded by face-saving and ICT questions suggests that they may alleviate turnout misreporting compared to the direct question. But do these techniques also yield inferences concerning the predictors of

---

<sup>9</sup> The complexity of CM designs can lead to noncompliance and misclassification, and thus less accurate measures of sensitive behaviors relative to a direct question (Höglinger and Diekmann 2017).

turnout that are more consistent with those drawn from data where misreporting is absent?

To address this question, we estimate demographic models of turnout for the 2015 British general election based on direct, face-saving and ICT questions. (Given its poor performance in estimating aggregate turnout, we do not estimate a model for the CM question.) We then compare each of these models against a benchmark model estimated using *validated* measures of individual turnout, based on official electoral records rather than respondent self-reports.<sup>10</sup> To the best of our knowledge, the only publicly available individual-level validated vote measures for the 2015 general election are those from the postelection face-to-face survey of the 2015 British Election Study (Fieldhouse, Green, Evans, Schmitt, Cees and Mellon 2016).<sup>11</sup> Generated via probability sampling and persistent recontact efforts, the 2015 BES face-to-face survey is widely considered to be the 'gold standard' among 2015 election surveys in terms of survey sample quality (Sturgis et al. 2016, 48). If the models estimated from online survey data using our turnout measures yield similar inferences to those estimated from the BES

---

<sup>10</sup> We must estimate distinct regression models for each question type because the ICT turnout measure does not yield individual-level turnout measures and therefore cannot be modelled using standard regression methods.

<sup>11</sup> Data from the online survey vote validation study reported in Rivers and Wells (2015) is not currently publicly available.

face-to-face data using validated turnout measures, we can be more confident that the former are properly correcting for misreporting.<sup>12</sup>

We estimate four regression models. First, a benchmark model is estimated using the 1,690 BES face-to-face respondents whose turnout was validated.<sup>13</sup> This is a binary logistic regression with a response variable coded as 1 if official records show a respondent voted and 0 otherwise. Our second and third models are binary logistic regressions estimated using our online survey data and have as their response variable turnout as measured by the direct question and the direct face-saving question,

---

<sup>12</sup> Note that differences between turnout models estimated from the two data sources may be due not only to residual misreporting in the online self-reports, but also to differences in the sample characteristics of a face-to-face versus an online survey. Indeed, Karp and Lühiste (2016) argue that turnout models estimated from online and face-to-face samples yield different inferences regarding the relationship between demographics and political participation. However, their evidence is based on direct and nonvalidated measures of turnout. It is possible that, once misreporting is addressed in both types of survey mode, inferences become more similar.

<sup>13</sup> Of this subsample, 1,286 ( 76.1%) voted. The 17 respondents who were measured as 'ineligible' to vote were coded as having not voted.

respectively. For our fourth model, we use the ICT regression methods developed in Imai (2011) to model the responses to the ICT question in our online survey.<sup>14</sup>

All four regression models include the same explanatory variables. First, we include a measure of self-reported party identification.<sup>15</sup> To avoid unduly small subsamples, respondents are classified into four groups: Conservative identifiers; Labour identifiers; identifiers of any other party; and those who do not identify with any party or who answer 'don't know'. Our second and third explanatory variables are age group (18-24; 25-39; 40-59; 60 and above) and gender (male or female). Our fourth explanatory variable is a respondent's highest level of educational qualification, classified according to the UK Regulated Qualifications Framework (no qualifications, unclassified qualifications or don't know; Levels 1-2; Level 3; Level 4 and above). These predictors constitute the full set of variables that are measured in a comparable format in both our experimental data and the 2015 BES face-to-face data.

Logistic regression coefficients are difficult to substantively interpret or compare across models. Therefore, we follow Blair and Imai (2012) and focus on predicted prevalence of the sensitive behavior for different political and demographic groups.

Specifically, for a given sample and regression model, we ask what the predicted turnout

---

<sup>14</sup> We estimate the ICT regression model using the 'list' package (Blair and Imai 2010) in R.

<sup>15</sup> For the online data this was measured by YouGov straight after the 2015 general election.

rate in the sample would be if all BES face-to-face respondents were assigned to a particular category on a variable of interest, while holding all other explanatory variables at their observed values.<sup>16</sup>

[INSERT FIGURE 2 ABOUT HERE]

Figure 2 graphs the group-specific predicted turnout rates for the regression models. The left panel shows that the regression based on the direct question (open circles) generates group-specific predicted turnout rates that all far exceed those from

---

<sup>16</sup> First, we simulate 10,000 Monte Carlo draws of the model parameters from a multivariate normal distribution with mean vector and variance-covariance matrix equal to the estimated coefficients and variance-covariance matrix of the regression model. Second, for each draw, we calculate predicted turnout probabilities for all respondents in the BES face-to-face sample - setting all respondents to be in the political or demographic group of interest and leaving other predictor variables at their actual value - and store the mean turnout probability in the sample. The result is 10,000 simulations of the predicted turnout rate if all respondents in the sample were in a particular category on a particular political or demographic variable, averaging over the sample distribution of the other explanatory variables. The point estimate for the predicted turnout rate is the mean of these 10,000 simulations, and the 95% confidence interval is given by the 2.5th and 97.5th percentiles. Our results are substantively unchanged if predicted turnout rates were calculated based on the experimental survey sample.



the benchmark validated vote model (filled circles).<sup>17</sup> It also performs poorly in terms of recovering how turnout is associated with most variables. While the benchmark model yields predicted turnout rates for older and more qualified voters that are noticeably higher than for younger and less qualified voters, there is barely any variation in turnout rates by age group and education according to the direct question model. Only with respect to party identification does the direct question model recover the key pattern present in the benchmark model: that those with no clear party identification are less likely to vote than those who do.

The middle panel shows that the regression based on the face-saving turnout question (open circles) improves somewhat on the direct question regression. The group-specific predicted turnout rates are generally slightly closer to the benchmark rates (filled circles), although most remain significantly higher. In terms of relative turnout patterns, there is some evidence of higher predicted turnout rates for higher age groups, but the differences between young and old voters are too small and predicted turnout rate barely varies by education. In addition, the difference in the predicted turnout rates of those with and without a clear party identity are actually more muted in the face-saving model than in the benchmark model or the direct question model.

---

<sup>17</sup> Supplementary Materials Section E graphs the corresponding differences in predicted turnout rates and Section D reports raw regression coefficients for each model.

The right panel shows that the regression based on the ICT turnout questions (open circles) improves on both the direct and face-saving models. Although the uncertainty surrounding each group-specific turnout rate is considerably greater, most point estimates are closely aligned with the benchmark rates (filled circles). Moreover, this is not simply the result of an intercept-shift: The ICT model also recovers relative patterns of turnout that are generally more consistent with the benchmark model. Regarding party identification, the difference in predicted turnout rates of those who do and do not have a clear party identification is of similar magnitude to that in the direct and face-saving models. Regarding age and education, as in the benchmark model, predicted turnout rates increase substantially with age-group and qualification level.<sup>18</sup> Predicted turnout for 18-24 year-olds seems unduly low. But there is considerable

---

<sup>18</sup> The differences between the group-specific predicted turnout rates from the ICT and direct models imply that younger voters and less qualified voters in particular tend to misreport voting. This is consistent with the differences between the BES benchmark model and the direct model in Figure 2 and with earlier UK vote validation studies. Swaddle and Heath (1989), for example, find that "the groups with the lowest turnout are the ones who are most likely to exaggerate their turnout". This is different from misreporting patterns found in US studies (Bernstein, Chadha, and Montjoy 2001).

uncertainty surrounding this estimate due to the small proportion of respondents in this age group in the online sample (see Appendix Table A.2).<sup>19</sup>

[INSERT TABLE 2 ABOUT HERE]

Table 2 summarizes the performance of the different models vis-a-vis the benchmark model. The first three columns show the mean, median and maximum absolute differences in predicted group-specific turnout rates across the 14 political and demographic groups listed in Figure 2, comparing the benchmark model with each of the three remaining models. According to all measures, the face-saving model performs slightly better than the direct question model. But the ICT model performs substantially better than both, reducing mean and median discrepancies from the benchmark model by almost two-thirds. The final column of Table 2 gives the fraction of group-specific predicted turnout rates that are significantly different from their benchmark model counterpart (0.05 significance level, 2-tailed).<sup>20</sup> While almost all of the predicted turnout rates from the direct and face-saving models are significantly different from their benchmark counterparts, this is the case for only 2 of the 14 predicted turnout rates from the ICT model.

---

<sup>19</sup> The confidence interval for this age group is also wide for the direct and face-saving models, but the uncertainty induced by small sample size is amplified by the inefficiency of the ICT measures.

<sup>20</sup> Significance tests are based on the Monte Carlo simulations described above.

Overall, this analysis suggests that, as well as generating better aggregate estimates of turnout, ICT outperforms other techniques when it comes to estimating how turnout varies across political and demographic groups.

## CONCLUSIONS

This paper compared the performance of several sensitive survey techniques designed to reduce turnout misreporting in postelection surveys. To do so, we ran an experiment shortly after the 2015 UK general election. One group of respondents received the standard BES turnout question. Another group received a face-saving turnout question previously untested in the UK. For a third group, we measured turnout using the crosswise model, the first time this has been tested in the turnout context. For a fourth group, we measured turnout using a new item-count question designed following current best practice.

ICT estimates of aggregate turnout were significantly closer to the official 2015 turnout rate. We also introduced a more nuanced approach to validating ICT turnout measures: comparing inferences from demographic models of turnout estimated using ICT measures to those from models estimated using validated vote measures. Inferences from the ICT model were consistently closer to, and often statistically indistinguishable from, those from the benchmark validated vote model.

Thus, in contrast to Holbrook and Krosnick (2010b) and Thomas et al. (2017), our findings suggest that carefully designed ICTs *can* significantly reduce turnout misreporting in online surveys. This suggests that, in settings where practical or financial

constraints make vote validation impossible, postelection surveys might usefully include ICT turnout questions.

We also found that the direct turnout question with face-saving options did improve on the standard direct question, both in the accuracy of aggregate turnout estimates and validity of demographic turnout models. However, consistent with previous research (e.g., Belli et al. 2006; Zeglovits and Kritzinger 2014), these improvements were moderate compared to those from ICT.

In contrast, CM performed no better or worse than the standard direct turnout question in terms of estimating aggregate turnout. Taken together with Holbrook and Krosnick (2010a), this finding highlights the difficulty of successfully implementing randomized response questions and variants thereof in self-administered surveys.

Of course, there are limitations to our findings. First, our evidence comes only from online surveys and the mechanisms behind social desirability bias may be different in this mode compared to when a respondent interacts with a human interviewer by telephone or face-to-face. That said, other studies do show that ICT reduces misreporting in telephone (Holbrook and Krosnick 2010b) and face-to-face surveys (Comşa and Postelnicu 2013). Second, a well-acknowledged drawback of ICT is its statistical inefficiency. While ICT significantly improves on other techniques *despite* this inefficiency, future research should investigate whether further efficiency-improving adaptations of the ICT design - such as the 'double-list experiment' (Droitcour et al. 1991) and combining direct questions with ICT (Aronow et al. 2015) - are effective in the context of turnout measurement. Finally, our regression validation focused only on

how basic descriptive respondent characteristics are correlated with turnout and our survey was conducted during one specific time period in relation to the election. Future research could also validate using attitudinal turnout correlates and could compare turnout questions when fielded closer to and further from Election Day.

## **Appendix: Information on survey samples**

### **Experimental survey data**

Our survey experiment was fielded via four online surveys run by YouGov. The fieldwork dates for each survey 'wave' were, respectively, 8-9 June (Wave 1), 9-10 June (Wave 2), 10-11 June (Wave 3) and 11-12 June 2015 (Wave 4). Table A.1 reports the sample size for each treatment group in each survey wave.

[INSERT TABLE A.1 ABOUT HERE]

The target population for each survey wave was the adult population of Great Britain. YouGov maintain an online panel of over 800,000 UK adults (recruited via their own website, advertising, and partnerships with other websites) and hold data on the sociodemographic characteristics and newspaper readership of each panel member. Drawing on this information, YouGov uses targeted quota sampling, not random probability sampling, to select a sub-sample of panelists for participation in each survey. Quotas are based on the distribution of age, gender, social grade, party identification, region and type of newspaper readership in the British adult population. YouGov has multiple surveys running at any time and uses a proprietary algorithm to determine, on a rolling basis, which panelists to email invites to and how to allocate invitees to surveys when they respond. Any given survey thus contains a reasonable number of panelists who are 'slow' to respond to invites. Along with the modest cash incentives YouGov offer to survey participants, this is designed to increase the rate at which less politically engaged panelists take part a survey.

Due to the way respondents are assigned to surveys YouGov do not calculate a per survey participation rate. However, the overall rate at which panelists invited to

participate in a survey do respond is 21%. The average response time for an email invite is 19 hours from the point of sending.

Descriptive statistics for the sample are provided in Table A.2 and a comparison to 2015 UK population characteristics is provided in Table A.3.

[INSERT TABLE A.2 ABOUT HERE]

[INSERT TABLE A.3 ABOUT HERE]

### **2015 British Election Study Face-to-Face Survey**

The 2015 British Election Study face-to-face study (Fieldhouse et al. 2016) was funded by the British Economic and Social Research Council (ESRC). Fieldwork was conducted by GfK Social Research between 8th May and 13th September 2015, with 97% of the interviews being conducted within three months of the general election date (7th May 2015). Interviews were carried out via computer assisted interviewing.

Full details of the sampling procedure are given in Moon and Bhaumik (2015). Here we provide a brief overview based on their account. The sample was designed to be representative of all British adults who were eligible to vote in the 2015 general election. It was selected via multistage cluster sampling as follows: first, a stratified random sample of 300 Parliamentary constituencies was drawn; second, two Lower Layer Super Output Area (LSOAs) per constituency were randomly selected, with probability proportional to size; third, household addresses were sampled randomly within each LSOA; fourth, one individual was randomly selected per household.

Overall, 2,987 interviews were conducted. According to the standard AAPOR conventions for reporting response rates this represents a 55.9% response rate



(response rate 3). Descriptive statistics for the sample are provided in Table A.2 and a comparison to 2015 UK population characteristics is provided in Table A.3.

Turnout was validated against the marked electoral register using the name and address information of face-to-face respondents who had given their permission for their voting behavior to be validated. The marked electoral register is the copy of the electoral register used by officials at polling stations on Election Day. Officials at polling stations put a mark on the register to indicate when a listed elector has voted. The marked registers are kept by UK Local Authorities for twelve months after Election Day. The BES team collaborated with the UK Electoral Commission, who asked Local Authorities to send copies of marked registers for inspection.<sup>21</sup> Respondents were coded into five categories based on inspection of the register (Mellon and Prosser 2015 Appendix B):

1. Voted: The respondents appeared on the electoral register and was marked as having voted.
2. Not voted - registered: The respondent appeared on the electoral register but was not marked as having voted.

---

<sup>21</sup> Despite persistent reminders from the BES team and their vote validation partner organisation, the Electoral Commission, several Local Authorities did not supply their marked electoral registers. As a result, overall the validated vote variable is missing for around 15% of the face-to-face respondents who agreed to be matched (Mellon and Prosser 2015).

3. Not voted - unregistered: The respondent did not appear on the electoral register but there was sufficient information to infer that they were not registered to vote, e.g., other people were registered to vote at the address or if no one was registered at the address people were registered at surrounding addresses.
4. Insufficient information: We did not have sufficient information in the register to assess whether the respondent was registered and voted, either because we were missing the necessary pages from the register or we had not been sent the register.
5. Ineligible: the respondent was on the electoral register but was marked ineligible to vote in the general election.

Mellon and Prosser (2015) report that validated turnout for a subset of respondents was coded by multiple coders, and that reliability was high (coders gave the same outcome in 94.8% of cases).

### **Supplementary Data**

Supplementary data are freely available at *Public Opinion Quarterly* online.

## References

- Abelson, Robert P., Elizabeth F. Loftus, and Anthony G. Greenwald. 1992. "Attempts to Improve the Accuracy of Self-Reports of Voting." In *Questions About Questions: Inquiries into the Cognitive Bases of Surveys*, edited by Judith M. Tanur, 138–53. New York: Russell Sage Foundation.
- Aronow, Peter, Alexander Coppock, Forrest W. Crawford, and Donald P. Green. 2015. "Combining List Experiments and Direct Question Estimates of Sensitive Behavior Prevalence." *Journal of Survey Statistics and Methodology* 3:43–66.
- Belli, Robert F., Sean E. Moore, and John VanHoewyk. 2006. "An Experimental Comparison of Question Forms Used to Reduce Vote Overreporting." *Electoral Studies* 25:751–59.
- Belli, Robert F., Michael W. Traugott, Margaret Young, and Katherine A. McGonagle. 1999. "Reducing Vote over-Reporting in Surveys: Social Desirability, Memory Failure, and Source Monitoring." *Public Opinion Quarterly* 63:90–108.
- Berent, Matthew K., Jon A. Krosnick, and Arthur Lupia. 2016. "Measuring Voter Registration and Turnout in Surveys: Do Official Government Records Yield More Accurate Assessments?" *Public Opinion Quarterly* 80:597–621.
- Bernstein, Robert, Anita Chadha, and Robert Montjoy. 2001. "Overreporting Voting: Why It Happens and Why It Matters." *Public Opinion Quarterly* 65:22–44.
- Blair, Graeme, and Kosuke Imai. 2010. "list: Statistical Methods for the Item Count Technique and List Experiment." Available at The Comprehensive R Archive Network (CRAN). <http://CRAN.R-project.org/package=list>.
- . 2012. "Statistical Analysis of List Experiments." *Political Analysis* 20:47–77.
- Blair, Graeme, Kosuke Imai, and Yang-Yang Zhou. 2015. "Design and Analysis of the Randomized Response Technique." *Journal of the American Statistical Association* 110:1304–19.
- Brehm, John. 1993. *The Phantom Respondents: Opinion Surveys and Political Representation*. Ann Arbor, MI: Michigan University Press.
- Bryan, Christopher J., Gregory M. Walton, Todd Rogers, and Carol S. Dweck. 2011. "Motivating Voter Turnout by Invoking the Self." *Proceedings of the National Academy of Sciences* 108: 12653–6.
- Cassel, Carol A. 2003. "Overreporting and Electoral Participation Research." *American Politics Research* 31:81–92.
- Clifford, Scott, and Jennifer Jerit. 2015. "Do Attempts to Improve Respondent Attention Increase Social Desirability Bias?" *Public Opinion Quarterly* 79:790–802.

Comşa, Mircea, and Camil Postelnicu. 2013. "Measuring Social Desirability Effects on Self-Reported Turnout Using the Item Count Technique." *International Journal of Public Opinion Research* 25:153–72.

Coutts, Elisabeth, and Ben Jann. 2011. "Sensitive Questions in Online Surveys: Experimental Results for the Randomized Response Technique (RRT) and the Unmatched Count Technique (Uct)." *Sociological Methods & Research* 40:169–93.

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati. 1991. "Measurement Errors in Surveys," edited by Paul B. Biemer, Robert M. Groves, Lars E. Lyberg, Nancy A. Mathiowetz, and Seymour Sudman 185–210. New York: Wiley.

Fieldhouse, Ed, Jane Green, Geoffrey Evans, Hermann Schmitt, van der Eijk Cees, Jonathan Mellon, and Chris Prosser. 2016. "British Election Study, 2015: Face-to-Face Postelection Survey. [Data Collection]." UK Data Service.

Fowler, Floyd J. 1995. *Improving Survey Questions: Design and Evaluation*. Thousand Oaks, CA: Sage Publications.

Glynn, Adam N. 2013. "What Can We Learn with Statistical Truth Serum? Design and Analysis of the List Experiment." *Public Opinion Quarterly* 77:159–77.

Hanmer, Michael J., Antoine J. Banks, and Ismail K. White. 2014. "Experiments to Reduce the over-Reporting of Voting: A Pipeline to the Truth." *Political Analysis* 22:130–41.

Hochstim, Joseph R. 1967. "A Critical Comparison of Three Strategies of Collecting Data from Households." *Journal of the American Statistical Association* 62:976–89.

Holbrook, Allyson L., and Jon A. Krosnick. 2010a. "Measuring Voter Turnout by Using the Randomized Response Technique: Evidence Calling into Question the Method's Validity." *Public Opinion Quarterly* 74:328–43.

———. 2010b. "Social Desirability Bias in Voter Turnout Reports: Tests Using the Item Count Techniques." *Public Opinion Quarterly* 74:37–67.

Holbrook, Allyson L., Melanie C. Green, and Jon A. Krosnick. 2003. "Telephone Vs. Face-to-Face Interviewing of National Probability Samples with Long Questionnaires: Comparisons of Respondent Satisficing and Social Desirability Bias." *Public Opinion Quarterly* 67:79–125.

Holtgraves, Thomas, James Eck, and Benjamin Lasky. 1997. "Face Management, Question Wording, and Social Desirability." *Journal of Applied Social Psychology* 27:1650–71.

Imai, Kosuke. 2011. "Multivariate Regression Analysis for the Outcome Count Technique." *Journal of the American Statistical Association* 106:407–16.

- Jackman, Simon. 1999. "Correcting Surveys for Non-Response and Measurement Error Using Auxiliary Information." *Electoral Studies* 18:7–27.
- Jann, Ben, Julia Jerke, and Ivar Krumpal. 2012. "Asking Sensitive Questions Using the Crosswise Model: An Experimental Survey Measuring Plagiarism." *Public Opinion Quarterly* 76:32–49.
- Jones, Edward E., and Harald Sigall. 1971. "The Bogus Pipeline: New Paradigm for Measuring Affect and Attitude." *Psychological Bulletin* 76:349–64.
- Jones, Emily. 2008. "Vote Overreporting: The Statistical and Policy Implications." *Policy Perspectives* 15:83–97.
- Karp, Jeffrey A., and David Brockington. 2005. "Social Desirability and Response Validity: A Comparative Analysis of Overreporting Voter Turnout in Five Countries." *Journal of Politics* 67:825–40.
- Karp, Jeffrey A., and Maarja Lühiste. 2016. "Explaining Political Engagement with Online Panels: Comparing the British and American Election Studies." *Public Opinion Quarterly* 80:666–93.
- Kish, Leslie. 1965. *Survey Sampling*. New York: John Wiley and Sons.
- Kuklinski, James H., Michael D. Cobb, and Martin Gilens. 1997. "Racial Attitudes and the 'New South'." *Journal of Politics* 59:323–49.
- Lax, Jeffrey R., Justin H. Phillips, and Alissa F. Stollwerk. 2016. "Are Survey Respondents Lying About Their Support for Same-Sex Marriage?" *Public Opinion Quarterly* 80:510–33.
- Locander, William, Seymour Sudman, and Norman Bradburn. 1976. "An Investigation of Interview Method, Threat and Response Distortion." *Journal of the American Statistical Association* 71:269–75.
- Lumley, Thomas. 2004. "Analysis of Complex Survey Samples." *Journal of Statistical Software* 9(8).
- McDonald, Michael P. 2003. "On the over-Report Bias of the National Election Study Turnout Rate." *Political Analysis* 11:180–86.
- Mellon, Jonathan, and Christopher Prosser. 2015. "Missing Non-Voters and Misweighted Samples: Explaining the 2015 Great British Polling Miss." BES Working Paper.
- Miller, Judith D. 1984. "A New Survey Technique for Studying Deviant Behavior."
- Moon, Nick, and Claire Bhaumik. 2015. "British Election Study 2015: Technical Report." GfK UK Social research.

- Persson, Mikael, and Maria Solevid. 2014. "Measuring Political Participation – Testing Social Desirability Bias in a Web-Survey Experiment." *International Journal of Public Opinion Research* 26:98–112.
- Presser, Stanley. 1990. "Can Context Changes Reduce Vote over-Reporting?" *Public Opinion Quarterly* 54:586–93.
- Rivers, Douglas, and Anthony Wells. 2015. "Polling Error in the 2015 Uk General Election: An Analysis of Yougov's Pre- and Postelection Polls." UK: YouGov Inc.
- Roese, Neal J., and David W. Jamieson. 1993. "Twenty Years of Bogus Pipeline Research: A Critical Review and Meta-Analysis." *Psychological Bulletin* 114:809–32.
- Rosenfeld, Bryn, Kosuke Imai, and Jacob N Shapiro. 2016. "An Empirical Validation Study of Popular Survey Methodologies for Sensitive Questions." *American Journal of Political Science* 60:783–802.
- Schenker, Nathaniel, and Jane F. Gentleman. 2001. "On Judging the Significance of Differences by Examining the Overlap Between Confidence Intervals." *The American Statistician* 55:182–86.
- Selb, Peter, and Simon Munzert. 2013. "Voter Overrepresentation, Vote Misreporting, and Turnout Bias in Postelection Surveys." *Electoral Studies* 32:186–96.
- Sturgis, Patrick, Nick Baker, Mario Callegaro, Stephe Fisher, Jane Green, Will Jennings, Jouni Kuha, Benjamin E. Lauderdale, and Patten Smith. 2016. *Report of the Inquiry into the 2015 British General Election Opinion Polls*. Market Research Society; British Polling Council.
- Swaddle, Kevin, and Anthony Heath. 1989. "Official and Reported Turnout in the British General Election of 1987." *British Journal of Political Science* 19:537–51.
- Tan, Ming T., Guo-Liang Tian, and Man-Lai Tang. 2009. "Sample Surveys with Sensitive Questions: A Nonrandomized Response Approach." *The American Statistician* 63:9–16.
- The Comparative Study of Electoral Systems. 2017. "CSES Module 4 [Data Collection]."
- Thomas, Kathrin, David Johann, Sylvia Kritzing, Carolina Plescia, and Eva Zeglovits. 2017. "Estimating Sensitive Behavior: The Ict and High Incidence Electoral Behavior." *International Journal of Public Opinion Research* 29:157–71.
- Tourangeau, Roger, and Ting Yan. 2007. "Sensitive Questions in Surveys." *Psychological Bulletin* 133:859–83.
- Voogt, Robert J. J., and Willem E. Saris. 2003. "To Participate or Not to Participate: The Link Between Survey Participation, Electoral Participation, and Political Interest." *Political Analysis* 11:164–79.

Warner, Stanley L. 1965. "Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias." *Journal of the American Statistical Association* 60:63–69.

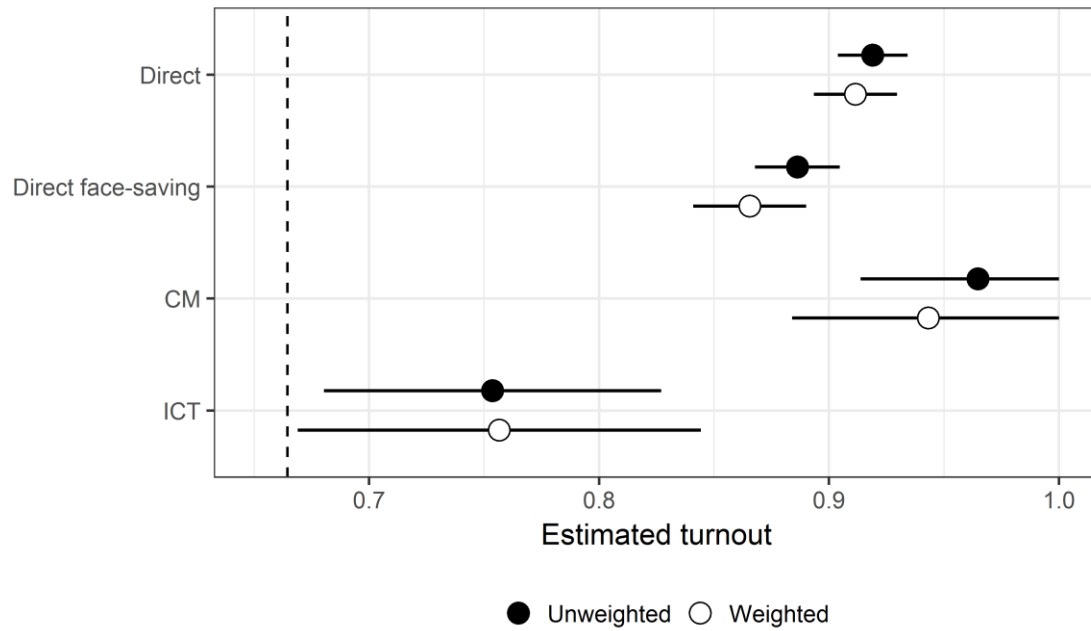
Wolter, Felix, and Bastian Laier. 2014. "The Effectiveness of the Item Count Technique in Eliciting Valid Answersto Sensitive Questions. an Evaluation in the Context of Self-Reported Delinquency." *Survey Research Methods* 8:153–68.

Yu, Jun-Wu, Guo-Liang Tian, and Man-Lai Tang. 2008. "Two New Models for Survey Sampling with Sensitive Characteristic: Design and Analysis." *Metrika* 67:251–63.

Zeglovits, Eva, and Sylvia Kritzinger. 2014. "New Attempts to Reduce Overreporting of Voter Turnout and Their Effects." *International Journal of Public Opinion Research* 26:224–34.

## Figures

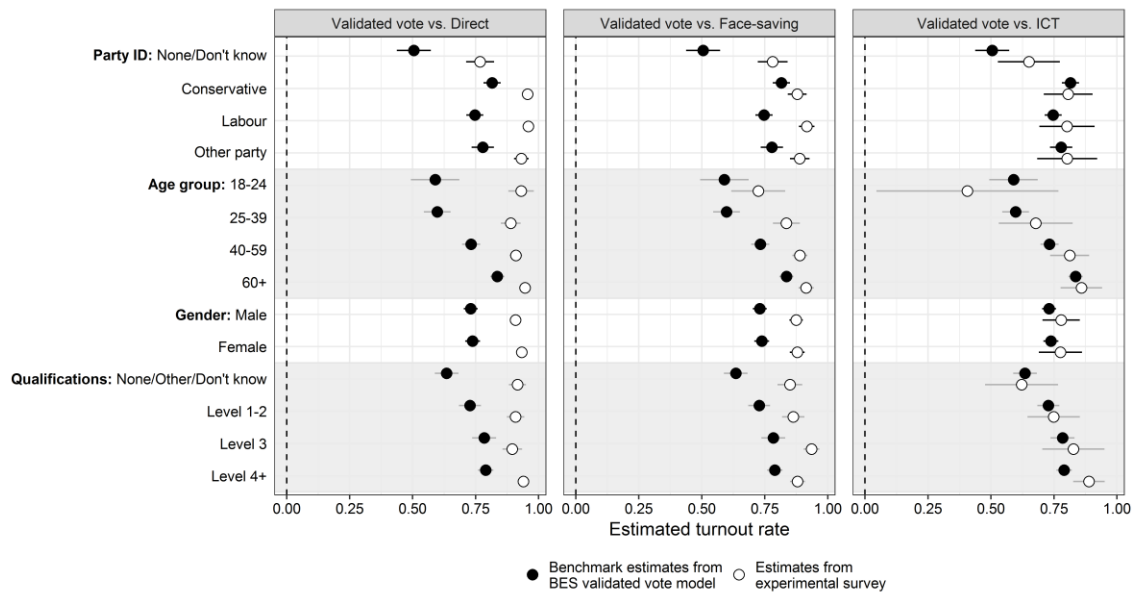
**Figure 1. 2015 Turnout Estimates by Estimation Method.**



Note: For each turnout question, points indicate weighted and unweighted point estimates for 2015 general election turnout. Lines indicate 95% confidence intervals. The dashed vertical line indicates actual GB turnout.



**Figure 2. Comparing Turnout Models Against BES Validated Vote Model.**



Note: Political and demographic groups are listed along the y-axis. For each group, we plot the predicted turnout rates based on a regression model, averaging over the distribution of covariates in the BES validated vote sample. Predicted turnout rates for the direct, face-saving and ICT regression models are shown, respectively, in the left, middle, and right panels (open circles). Predicted turnout rates from the benchmark BES validated vote model are displayed in every panel (filled circles). Reading example: the filled circle for 'Male' in each panel indicates that, based on the validated vote regression model, if we set all respondents in the BES sample to 'Male' while holding all other explanatory variables at their observed values, the predicted turnout rate would be 73.8% [70.1%, 76.7%].

## Tables

**Table 1. Rates of 'Don't know' Responses by Treatment Group.**

<b>Method</b>	<b>'Don't know' rate</b>
Direct	0.020
Face-saving	0.016
ICT control	0.036
ICT sensitive	0.021
CM	0.036

Note: Entries show, for each treatment group, the rate of 'don't know' responses to the item measuring turnout.

**Table 2. Summary of differences from benchmark validated vote model.**

<b>Method</b>	<b>Absolute differences</b>			<b>Sig. Difference</b>
	<b>Mean</b>	<b>Median</b>	<b>Maximum</b>	
Direct	0.2	0.18	0.34	14/14
Face-saving	0.15	0.14	0.28	13/14
ICT	0.06	0.05	0.18	2/14

Note: For a given test model (row), the first three columns show the mean, median and maximum absolute discrepancy between group-specific predicted turnout rates generated by this model and those generated by the benchmark validated vote model. The final column gives the fraction of group-specific predicted turnout rates that are significantly different from their benchmark model counterpart (0.05 significance level, 2-tailed).

**Table A.1. Treatment sample sizes by survey wave.** This table shows the distribution of treatment assignment by survey wave.

Wave	Treatment Group					All
	Direct	Face-Saving	ICT Control	ICT Sensitive	CM	
1	307	289	292	333	295	1516
2	335	312	350	342	311	1650
3	326	271	329	290	314	1530
4	292	281	324	321	314	1532

**Table A.2. Sample characteristics: experimental data versus 2015 BES face-to-face**

**survey.** All respondent attributes were coded as binary indicators. Columns 1-2 and 3-4 summarize, respectively, the distribution of each indicator in our experimental data and in the 2015 BES face-to-face sample.

	<b>Experimental</b>		<b>BES</b>	
	<b>N</b>	<b>Mean</b>	<b>N</b>	<b>Mean</b>
Age group 18-24	6227	0.08	2955	0.07
Age group 25-39	6227	0.18	2955	0.21
Age group 40-59	6227	0.41	2955	0.35
Age group 60+	6227	0.33	2955	0.37
Female	6228	0.52	2987	0.54
Male	6228	0.48	2987	0.46
Qualifications None/Other/Don't know	6228	0.21	2987	0.30
Qualifications Level 1-2	6228	0.22	2987	0.21
Qualifications Level 3	6228	0.19	2987	0.13
Qualifications Level 4+	6228	0.38	2987	0.36
Party ID None/Don't know	6228	0.17	2964	0.16
Party ID Conservative	6228	0.29	2964	0.31
Party ID Labour	6228	0.29	2964	0.32
Party ID other party	6228	0.24	2964	0.21
Social grade DE	6228	0.20		
Social grade C2	6228	0.15		
Social grade C1	6228	0.26		
Social grade AB	6228	0.39		
Wave 1	6228	0.24		
Wave 2	6228	0.26		
Wave 3	6228	0.25		
Wave 4	6228	0.25		
Direct treatment	6228	0.20		
Face-saving treatment	6228	0.19		
ICT control treatment	6228	0.21		
ICT sensitive treatment	6228	0.21		
CM treatment	6228	0.20		

**Table A.3. Sample characteristics compared to 2011 Census.** The first two column show the relative frequency of age groups and gender in the experimental data and in the 2015 BES face-to-face survey. The final column shows the GB population frequency of each demographic group according to the 2011 Census.

	<b>Experimental</b>	<b>BES</b>	<b>Census</b>
Age group 18-24	8.0	7.3	11.8
Age group 25-39	18.3	20.9	25.4
Age group 40-59	41.1	34.9	34.2
Age group 60+	32.6	36.9	28.6
Female	52.3	54.1	51.4
Male	47.7	45.9	48.6